Estimating Accident Potential of Ontario Road Sections

BHAGWANT N. PERSAUD

The identification of unsafe road locations (blackspots) and the evaluation of treatment effectiveness should be based on the number of accidents expected in the long run (accident potential) rather than on the short-term count. A method for estimating the underlying accident potential of Ontario road sections, using accident and other data, is presented. The method first uses regression models to produce an initial estimate of a section's accident potential on the basis of its traffic and geometric characteristics. This estimate is then refined by being combined with the section's accident count, using an empirical Bayesian procedure. The results indicate that the empirical Bayesian estimates are superior to those based on the accident count or the regression prediction by themselves, particularly for sections that might be of interest in a program to identify and treat unsafe road locations.

The efficient allocation of resources to highway safety programs requires that accident blackspots (unsafe intersections, road sections, etc.) be properly identified and that estimates of the safety benefit of a potential treatment be as sound as possible. Crucial to both aspects is the proper estimation of accident potential of a site being considered for treatment.

For blackspot identification, it has been common to use variants of the "rate-and-number" method (*1*), in which sites registering an unusually high accident potential are selected for detailed examination.

For safety benefits, the safety effect of a treatment is usually estimated from its previous applications by comparing the accident potential of a site before and after it was treated. This estimate of safety effect is then applied to the current estimate of accident potential of a site under consideration for the treatment.

In both cases, recent research (*2,3*) has recognized that accident potential should be the average number of accidents expected in the long run on a site, not the short-term count. For obvious reasons, this underlying long-term accident potential cannot be observed, so it must be estimated.

The primary focus of the research on which this paper is based was to provide estimates of accident potential for Ontario road sections, using data readily available to safety analysts at the Ontario Ministry of Transportation (MTO). The work described is an application of a recently developed empirical Bayesian approach (*3*) that refines prior estimates of accident potential obtained from multivariate regression analysis.

Department of Civil Engineering, Ryerson Polytechnic Institute, 350 Victoria Street, Toronto M5B 2K3 Canada.

DATA

Raw computer data were obtained from two sources at MTO for roads under the ministry's jurisdiction. Inventory information for Linear Highway Referencing System (LHRS) sections and subsections contained information on highway division, category and environment, section length, lane and shoulder widths, roadway and shoulder surface type, number of lanes, speeds and speed limits, and other geometric characteristics.

The other raw data set contained, for each of the years 1983–1986, the total number of accidents and traffic information for LHRS sections and subsections. The traffic information included percentage of commercial vehicles, seasonal and annual average daily traffic (ADT) volumes, and directional split. Details of the accidents were not available, but the absence of more detailed data should not be seen as a shortcoming because the aim of the research was to build functional models that use readily available data.

Models were desired for each of three road clasess. Final data sets were prepared for each road class by combining the raw traffic and inventory data for as many LHRS sections or subsections as possible. Summary information for each final data set is shown in Table 1.

The summary figures in Table 1 show that from 1985 to 1986, Class 1 roads experienced a 20 percent increase in accidents and from 1983 to 1984, a 12 percent increase in traffic. Comparable figures for the other classes show that Class 2 roads had a 7 percent increase in accidents and a 4 percent increase in traffic, while Class 3 roads had a 4 percent increase in accidents and a 3 percent increase in traffic.

METHODOLOGY

The procedure for developing and applying the models for estimating accident potential of road sections closely follows that for estimating accident potential for rail-highway grade crossings (*4*), signalized intersections (*5*), and Ontario drivers (*6*). Further details of the methodology can be obtained from these papers as well as from a recent FHWA publication (*3*). This section summarizes the bare essentials of the methodology.

The procedure provides an estimate of the accident potential of a road section, given (a) its accident history and (b) its traffic, geometric, and other characteristics. The method used to reach this estimate is, in effect, an empirical Bayesian procedure (*7*) that mixes the two sets of information. This procedure is depicted in Figure 1 and summarized in the following as a sequence of two steps.

TABLE 1 SUMMARY INFORMATION FOR FINAL DATA SETS

	CLASS 1 Freeways	CLASS 2 Other Primary	CLASS 3 Secondary Tertiary
Number of sections	404	1680	166
Total length (km)	1583.6	12024.6	2396.5
Average daily traffic (Weighted by length)			
1983	24298	2548	318
1984	25129	2608	326
1985	26219	2610	324
1986	28981	2767	338
Total Accidents			
1983	11612	15853	564
1984	12533	16963	601
1985	14641	18125	662
1986	14420	17033	554

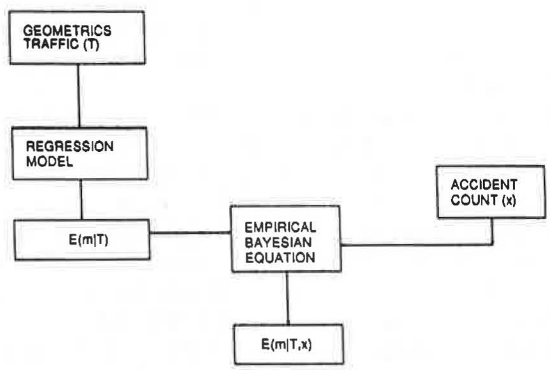

FIGURE 1 Procedure for estimating accident potential.

Regression Models for Initial Prediction of Accident Potential (Step 1)

Generalized linear modeling using the GLIM computer package (*8*) was used to estimate $E(m|T)$, the underlying accident potential of a section for the period 1983–1984, given its characteristics (e.g., traffic volume) for this period. The model form used was

$$E(m|T) = \text{SCL} * a1 * (\text{ADT})^{b1} \quad (1)$$

where

T = set of traffic and geometric characteristics,
SCL = section length (km),
$a1$ and $b1$ = model parameters estimated by GLIM.

The observed accident count on a section in the period 1983–1984 was used as an estimate of the dependent variable $E(m|T)$. GLIM allows the specification of a negative binomial error structure for the dependent variable. This is now known to be more appropriate for accident counts than the traditional normal distribution (*3,5,9*).

The negative binomial error specification follows on assumptions that $(m|T)$ is gamma distributed and that accident occurrence on a section follows the Poisson probability law. Under these conditions (*3,5,9*), the variance of the regression estimates can be estimated from

$$\text{Var}(m|T) = [E(m|T)]^2/k \quad (2)$$

where k is a parameter of the negative binomial distribution.

The procedure for estimating k is iterative. It is first necessary to specify an initial guess of the value of k in order to calibrate the regression model that estimates $E(m|T)$. This guess is then refined by comparing it with the value of k estimated from a maximum likelihood procedure that assumes that each squared residual of the regression model is an estimate of $\text{Var}(m|T)$ and that each count comes from a negative binomial distribution with mean $E(m|T)$ and variance given by Equation 2.

Rearranging the terms of Equation 2 indicates that k can be used as a measure of the variation explained by the regression model, that is, the larger the value of k, the more variation is explained. This is useful because the more traditional R^2 measure is not appropriate when GLIM is used with a negative binomial structure.

Empirical Bayesian Revised Estimates of Accident Potential (Step 2)

In general, two road sections that are similar in all of the independent variables used in the regression model will still be different in true accident potential even though they have the same model predictions according to Equation 1. This is because it is not possible to account in the regression model for all the factors that cause differences in accident potential (e.g., weather).

To account for this shortfall—in effect, to reduce the variation not explained by the regression model—$E(m|T)$ from Equation 1 can be further refined for an individual road section using the accident count (x) on that section to give $E(m|T, x)$, a final revised estimate of accident potential.

It can be shown (*3*) that, under the assumptions stated earlier (i.e., that the variation in $m|T$ can be described by a gamma probability distribution and that accident occurrence on a section follows the Poisson probability law), the revised estimate of accident potential is

$$E(m|T,x) = wE(m|T) + (1 - w)x \quad (3)$$

where

$$w = [1 + E(m|T)/k]^{-1} \quad (4)$$

in which, from Equation 2, $k = E(m|T)^2/\mathrm{Var}(m|T)$.

It can also be shown (*3*) that

$$\frac{\mathrm{Var}(m|T,x) = E(m|T,x)}{1 + [E(m/T)/\mathrm{Var}(m|T)]} \quad (5)$$

The value obtained from Equation 3 is known as an empirical Bayesian estimate. It turns out that the variation in $\{m|T,x\}$ can also be described by a gamma distribution.

Before moving on, it is instructive to explore the meaning of Equation 3. This equation shows that $E(m|T,x)$, the estimated accident potential of a section, is a mixture of what is observed (x) and of $E(m|T)$—what is predicted on the basis of its characteristics (ADT, etc.).

If the prediction has a large variance, that is, there is much unexplained variation, then $\mathrm{Var}(m|T) >> E(m|T)$, w would be small and the accident potential would be close to x. This situation typically arises, according to the hypothesized model, for sections that are long or have a high traffic volume. For such sections, in effect, the accident count could reasonably be used as an estimate of accident potential and the benefit of empirical Bayesian estimation would be marginal.

Conversely, for $\mathrm{Var}(m|T) << E(m|T)$, the accident potential would be close to $E(m|T)$ and a relatively small weight would be given to the observed count (x).

ANALYSIS AND RESULTS

This section is divided into two parts. First, the development of regression models of accident potential for Ontario road sections is described. Next, the empirical Bayesian estimation of accident potential is illustrated.

Regression Models for Initial Estimation of Accident Potential

In this section, the development of the regression models is described. It should be emphasized that these models should not be judged on their ability to explain the causal factors related to accident occurrence. In particular, they should not be used to answer "what if" questions about the impact of altering values of the independent variables.

Equation 1 indicates that the models are of the form

$$E(m|T) = \mathrm{SCL} * a1 * (\mathrm{ADT})^{b1}$$

where $\mathrm{Var}(m|T) = E(m|T)^2/k$. This form is quite common because it ensures that predicted accidents would be zero for an ADT of zero, but it does not, a priori, assume a linear relationship between accidents and traffic volume (*9*).

Whereas GLIM allows the error structure to be specified for $E(m|T)$, it can actually estimate a model for some specified "link" function of $E(m|T)$. In this case GLIM actually estimated models of the log (base e) form

$$\log[E(m|T)] = \log(\mathrm{SCL}) + \log(a1) + b1 * \log(\mathrm{ADT}) \quad (6)$$

In GLIM, the term log(SCL) is specified as an offset that is subtracted from each point estimate of $\log[E(m|T)]$.

The geometric variables selected for the regression models varied with class of road, which should not be surprising because the classes differ significantly in traffic and geometric characteristics. For each class the variable selection was accomplished by using similar statistical procedures to those detailed elsewhere (*8,9*). Models for each road class are described separately later.

Class 1

Freeway sections tend to be all multilane and divided, have the same high geometric standards, and be similar in other such features as speed limit, so it appeared unlikely that a model could incorporate geometric variables and be significantly better than the traffic volume model for this class. Indeed, it turned out that all attempts to add geometric variables—most notably, the number of lanes—were unsuccessful.

The estimates of $a1$ and $b1$ produced by GLIM are given in Table 2 along with other relevant details.

Class 2

The Class 2 road sample was seen as a mixture of urban and rural roads, two-lane and multilane roads, and divided and undivided roadways, giving six possible categories (not eight, because there are no two-lane divided highways). Pertinent details of the road sections in each category are shown in Table 3.

Exploratory analysis revealed that both speed and speed limit varied significantly from category to category but that within each category the variation was quite small. Similarly, for each geometric characteristic that might affect accident occurrence, the variation within individual categories was small. Thus, the use of a road section's category as a variable in an accident prediction model was found to be sufficient to account for variation attributable to geometric and speed factors. The final models for Class 2 roads reflect this reasoning. All attempts at incorporating other variables failed.

The final model form used the full Class 2 data set and allowed the ADT coefficient ($b1$) to vary with the two categorical variables, two-lane/multilane and rural/urban, whereas

TABLE 2 REGRESSION MODEL FOR CLASS 1 ROADS

Model Parameter	Estimated Value
a1 for ADT in thousands	0.6278
b1	1.024
Other statistics: k = 2.95; Observations = 404	

TABLE 3 CATEGORIES FOR CLASS 2 ROADS

	RURAL			URBAN*		
	2-lane	Multilane Divided	Multilane Undivided	2-lane	Multilane Divided	Multilane Undivided
Sections	1400	49	90	73	8	60
Length,km	11413.1	188.7	219.9	106.3	5.3	91.3
Average ADT						
1983	2175	8459	11379	4683	16551	12396
1984	2218	8575	11886	4826	18074	13146
1985	2206	8631	12371	4925	19923	13435
1986	2337	9240	13152	5185	20891	14206
Accidents						
1983	12273	570	1429	369	44	1168
1984	13000	724	1631	370	53	1185
1985	13879	740	1660	406	74	1370
1986	12735	653	1767	400	102	1375

* -- The urban category includes all roads with MTO road environment codes 2 and 3 (semi-urban and urban).

the coefficient $a1$ also varied with the categorical variable divided/undivided.

GLIM provided estimates of $\log(a1)$ and $b1$ in Equation 6 for a base category (rural, undivided, two-lane) along with adjustments to be applied to these values for the other five categories. It turns out that the estimated coefficients obtained in this way are the same as those that would have been obtained by separately estimating Level 1 traffic volume models for each category. The estimated values of $a1$ and $b1$ for each category are shown in Table 4.

TABLE 4 REGRESSION MODELS FOR CLASS 2 ROADS

Model Parameter	Estimated value
a1 for ADT in thousands for:	
Rural/undivided/2-lane	1.3392
Rural/undivided/multilane	0.6528
Urban/undivided/2-lane	3.6514
Urban/undivided/multilane	1.4196
Rural/divided/multilane	0.4591
Urban/divided/multilane	0.9984
b1 for:	
Rural/2-lane	0.8310
Rural/multilane	1.3037
Urban/2-lane	0.5588
Urban/multilane	0.8763

Other statistics: k=2.90; Observations=1680

Again, the estimated coefficients are for a 1-km section. Thus, the accident potential of a rural, undivided, multilane road section of length SCL km, for example, could be estimated from

$$E(m|T) = \text{SCL} * 1.3392 * \text{ADT}^{0.8310}$$

Class 3

The modeling exercise for Class 3 roads was similar to that done for Class 2 roads except that now the best explanatory variables turned out to be surface width and surface type. Additional explanatory variables provided little or no improvement.

Exploratory analysis resulted in selection of surface width and surface type as categorical variables, each with two levels, giving four category combinations. Table 5 describes these categories and provides relevant data for each category.

The final model form used the full Class 3 data set and allowed the ADT coefficient ($b1$) and the constant term [$\log(a1)$] to vary with the two categorical variables narrow/wide pavement and low/high class surface. GLIM provided estimates of $\log(a1)$ and $b1$ in Equation 6 for a base category (pavement width $\leq$ 6.1 m, surface type $\leq$ 4) and adjustments to be applied to these values for the other three categories. The

TABLE 5 CATEGORIES FOR CLASS 3 ROADS

	Pavement width ≤ 6.1 m		Pavement width >6.1 m	
	High Class Surface*	Lower Class Surface*	High Class Surface	Lower Class Surface
Sections	11	75	55	25
Length,km	101.2	999.1	771.5	524.7
Average ADT				
1983	625	219	455	246
1984	607	220	477	247
1985	613	220	463	259
1986	637	229	484	271
Accidents				
1983	43	167	257	97
1984	34	186	259	122
1985	41	190	302	129
1986	43	162	248	100

* -- High Class surfaces have MTO surface type codes >4 and consist of concrete, asphalt on concrete pavements and high class bituminous pavements. Lower class surface types have codes ≤ 4.

estimated values of $a1$ and $b1$ for each category are shown in Table 6.

As before, the estimated coefficients are for a 1-km section. Thus, the accident potential of a narrow Class 3 road section with a high class surface and length of SCL km, for example, could be estimated from

$$E(m|T) = \text{SCL} * 0.8988 * \text{ADT}^{0.3884}$$

Illustrating the Method

Consider again a narrow Class 3 road section with a high class surface. Suppose the section is 0.6 km long, has an ADT of 5,900, and recorded six accidents during 1983–1984.

From this example, the relevant regression model gives

$$E(m|T) = 0.6 * 0.8988 * 5.9^{0.3884} = 1.075$$

From Table 6, $k = 2.81$, giving $\text{Var}(m|T) = 1.075^2/2.81 = 0.416$. From Equation 4, $w = 0.72$, which gives, from Equation 3, $E(m|T,x) = 2.46$.

This is the value of accident potential that should be used in the blackspot identification process and in estimating the effect of any treatment that might be applied. The variance of this estimate, $\text{Var}(m|T,x)$, is 0.30 (from Equation 5). Without a model, the best estimate of accident potential is 6, which, it appears, might have been a randomly high accident count.

VALIDATION

It was shown earlier how an initial accident potential estimate from a regression model could be combined with the actual accident count on a road section to yield an empirical Bayesian estimate of accident potential for that section.

The point was made, however, that this estimation process (as opposed to using the accident count as an estimate of accident potential) is not as vital for sections with higher values of $E(m|T)$—the regression model prediction. Thus, it is reasonable that the true test of the method should be in its application to sections with low values of $E(m|T)$, that is, those that are short or have low traffic volume. Because the data obtained were at the level of LHRS sections, many of which tend to be relatively long or have high values of $E(m|T)$, there could be only limited testing of the empirical Bayesian procedure with the current data set.

TABLE 6 REGRESSION MODELS FOR CLASS 3 ROADS

Model Parameter	Estimated value
log(a1) for ADT in thousands for:	
Narrow/low class	0.9961
Narrow/high class	0.8988
Wide/low class	1.2348
Wide/high class	1.3336
b1 for:	
Narrow/low class	0.5844
Narrow/high class	0.3884
Wide/low class	0.7688
Wide/high class	0.6313

Other statistics: k = 2.81; Observations=166

Nevertheless it was possible to demonstrate the usefulness of the models using a sample of Class 2 roads. The sample consists of all 505 rural, two-lane road sections in the data set that are no more than 4 km long. From this sample, the road section having the highest value of accidents per kilometer during 1983–1984 in each of several ADT ranges was identified as a potential blackspot (the width of the range varied to allow for 10 to 15 sections in each range). From the model forms indicated by Equation 1 and Table 2, it can be seen that, because ADT is now the only regression variable in this sample, the same sections would be identified if the selection were properly based on $E(m|T,x)$ per kilometer.

The 39 sections identified are listed in Table 7 along with other items, including the ADT, accidents, regression model predictions, and empirical Bayesian estimates of accident prediction. Period 1 is 1983–1984; Period 2 is 1985–1986.

Several observations follow from the results in Table 7. First, there is a substantial decrease in accident count from Period 1 to Period 2. Assuming that the sections were largely untreated during those 4 years (a reasonable assumption, according to information in the inventory file), the logical conclusion is that this difference occurs because of a random up-fluctuation in the accident count during Period 1. The difference is even more pronounced if the Period 1 count is adjusted to account for the fact that rural Class 2 roads had a 5.3 percent increase in accidents in Period 2.

The second observation is that, because the sections were not selected on the basis of the Period 2 accident count, this count is an unbiased estimate of true accident potential of these sections. On this basis, an unbiased estimate of the total of the true accident potentials of these sections is 503. Thus, using the Period 1 accidents as estimates of accident potential overestimates this value by 34 percent. By contrast, using the regression prediction by itself as an estimate of accident potential underestimates this value (compare 243 and 503).

The third and most important observation is the close correspondence between the total of the empirical Bayesian estimates, $E(m|T,x)$ (487, or the adjusted value of 512), and the total of the estimates of true accident potential (503).

Finally, note that, even though the Period 2 accident count is an unbiased estimate of the true accident potential, the variance of this estimator (equal to the count, assuming a Poisson distribution of the total count on these sections) is substantially higher than that of the empirical Bayesian estimator (shown in the last column of Table 7). This is also the case when the variance of the regression prediction is compared with the variance of the empirical Bayesian estimator.

Though the final observation is of considerable interest, it is immaterial when considering the practical need to use present accident counts to estimate *future* accident potential of road sections. In such an application, the only alternative estimators are the recent accident count, the emperical Bayesian estimator, and perhaps the regression prediction. On the basis of this limited validation exercise, it is apparent that the empirical Bayesian estimator might be preferred.

SUMMARY AND CONCLUSIONS

The research used data available at the level of total accidents on relatively long road sections. Despite this limitation, useful

TABLE 7 ACCIDENT POTENTIAL OF SELECTED RURAL, TWO-LANE CLASS 2 ROADS

	DATA FOR PERIODS 1, 2									
	Section km		AADT		Accidents			Var(m¦T)		Var(m¦T,x,n)
LHRS#	Start	Length	1	2	1	2	E(m¦T)		E(m¦T,x,n)	
26420	0.0	1.2	195	210	8	5	0.413	0.059	1.359	0.169
43250	0.0	1.3	650	655	12	4	1.217	0.511	4.405	1.302
18170	10.3	1.0	665	655	11	8	0.954	0.314	3.441	0.852
34540	0.7	0.7	1050	1050	10	5	0.976	0.329	3.249	0.818
25080	4.2	1.6	1200	1275	6	6	2.493	2.144	4.114	1.902
10375	0.0	1.3	1375	1475	3	0	2.268	1.774	2.590	1.137
25483	0.0	2.9	1400	1550	8	7	5.137	9.099	6.967	4.453
26810	0.0	2.1	1625	1600	33	24	4.210	6.112	21.258	12.587
21600	1.1	1.1	1675	1750	61	33	2.262	1.764	27.998	12.267
43800	0.0	3.5	1725	2150	15	8	7.374	18.750	12.847	9.221
17400	0.0	1.0	1875	1950	5	4	2.258	1.758	3.458	1.514
11900	0.0	1.6	2000	2000	8	8	3.812	5.010	6.190	3.516
16940	4.4	2.2	2200	2100	11	8	5.673	11.099	9.198	6.087
10360	0.0	1.1	2350	2450	7	4	2.996	3.096	5.031	2.557
11330	0.0	0.6	2500	2550	4	4	1.721	1.021	2.569	0.957
46250	0.0	2.2	2600	2625	16	12	6.518	14.650	13.080	9.053
10350	7.2	1.2	2750	2925	8	6	3.725	4.785	6.129	3.446
32410	5.4	2.0	2800	2925	9	8	6.302	13.695	8.150	5.581
20040	0.0	1.0	2950	3100	10	13	3.291	3.734	6.857	3.645
19410	0.0	0.7	3050	3325	8	9	2.368	1.934	4.900	2.203
14946	0.0	2.1	3225	3775	12	9	7.442	19.096	10.722	7.715
10770	8.4	2.1	3425	3550	12	12	7.823	21.104	10.870	7.931
20715	0.0	2.0	3625	2800	20	12	7.810	21.035	16.699	12.178
14680	0.0	1.2	3700	3725	10	5	4.767	7.835	8.020	4.987
34735	2.7	1.2	3850	6350	10	7	4.927	8.370	8.120	5.111
35240	6.3	1.1	4125	3975	11	9	4.783	7.887	8.653	5.387
23460	0.0	1.4	4225	3575	12	6	6.209	13.295	10.157	6.923
16140	0.0	1.6	4475	5450	1	11	7.444	19.106	10.003	7.198
19435	0.0	2.3	4600	4800	28	21	10.948	41.331	24.429	19.313
23706	4.7	1.9	4900	4550	24	5	9.532	31.328	20.625	15.814
12200	3.9	4.0	5250	5400	37	44	21.251	155.720	35.109	30.893
29500	1.9	0.4	5600	5850	6	0	2.242	1.734	3.881	1.692
28540	0.0	0.6	5900	6050	17	12	3.512	4.254	10.900	5.970
33010	1.8	0.8	6050	6300	8	7	4.782	7.885	6.785	4.224
40330	2.8	1.6	6850	6650	34	27	10.603	38.768	28.975	22.752
35430	0.0	1.5	7525	7150	23	23	10.748	39.834	20.397	16.063
16490	1.9	1.4	8500	8950	26	33	11.100	42.488	22.914	18.167
16100	0.0	2.2	10200	10400	49	44	20.297	142.055	45.412	39.734
10550	0.0	0.8	12450	13650	38	30	8.710	26.162	30.684	23.020
TOTAL		60.5			641	503	230.9	750.9	487.1	338.3
(ADJUSTED)*					**(675)**		**(243)**		**(513)**	**(375)**

* **Adjustments account for a 5.3% increase in accidents from period 1 to 2.**

insights were gained into the estimation of accident potential for Ontario road sections. A summary and important conclusions of the research follow:

1. An empirical Bayesian procedure for the estimation of accident potential of road sections was presented. This estimator combines the recent accident count with a regression model prediction of accidents expected on the basis of the section's traffic volume and geometric characteristics.
2. The regression models developed make sense intuitively, but, because they are intended for use in initial accident prediction from variables whose values are readily available, they should not be judged on their ability to explain the causal factors related to accident occurrence. In particular, they should not be used to answer "what if" questions about the impact of altering values of the independent variables.
3. When the regression predictions were combined with the accident count information for a set of potential blackspot sections, using the resulting empirical Bayesian estimates was observed to be superior to using the accident count or the regression prediction alone. The results confirm that the accident count of a road section is, by itself, not a good estimator of accident potential, particularly for road sections that might be of interest when considering and applying treatment in a blackspot identification and treatment process. The accident count is a reasonable estimator of accident potential if sections are identified at random, that is, without regard to the accident count itself, and then only, according to the hypothesized models, for sections with that have relatively high traffic volume, section length, or accident frequency.
4. The procedures presented can be used for estimating accident potential for a number of applications, including (a) estimating the safety effect of improvements, (b) identifying potential blackspots, and (c) estimating the safety benefits (or disbenefits) of potential treatments at blackspots or other sites being considered in a resource-allocation procedure.
5. The procedures presented are currently useful for Ontario road sections, but it is still desirable to test their applicability outside Ontario and to seek to build improved regression models of accident potential. For example, separate models could be constructed for single-vehicle accidents and collisions, and for intersection and nonintersection accidents. Sim-

ilar models should be constructed for roads and intersections under the jurisdiction of municipalities.

ACKNOWLEDGMENTS

This paper resulted from research performed under a contract with the research and development branch of MTO. This support, along with the guidance of Alex Ugge, is gratefully acknowledged.

REFERENCES

1. J. I. Taylor and H. T. Thompson. Identification of Hazardous Locations. Report FHWA-RD-77-83. FHWA, U.S. Department of Transportation, 1977.
2. J. L. Higle and J. M. Witkowski. Bayesian Identification of Hazardous Locations. In *Transportation Research Record 1185*, TRB, National Research Council, Washington, D.C., 1988, pp. 24–36.
3. E. Hauer. Empirical Bayesian Approach to the Estimation of Unsafety—The Multivariate Approach. Report FHWA-RD-90-006. FHWA, U.S. Department of Transportation, 1990.
4. E. Hauer and B. N. Persaud. How To Estimate the Safety of Rail-Highway Grade Crossings and the Safety Effects of Warning Devices. In *Transportation Research Record 1114*, TRB, National Research Council, Washington, D.C., 1987, pp. 131–140.
5. E. Hauer, J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185*, TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
6. E. Hauer, B. N. Persaud, A. Smiley, and D. Duncan. Estimating the Accident Potential of Ontario Drivers. *Accident Analysis and Prevention* (in press).
7. C. N. Morris. Parametric Empirical Bayesian Inference: Theory and Applications. *Journal of the American Statistical Association*, Vol. 78, No. 3, 1983, pp. 47–65.
8. R. J. Baker and J. A. Nelder. *Generalized Linear Interactive Modelling. The GLIM System—Release 3.* Rothamsted Experimental Station, Harpenden, Hertfordshire, England, 1978.
9. G. Maycock and R. D. Hall. *Accidents at 4-Arm Roundabouts.* Report 1120. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.